GUSTAF ARRHENIUS

CAN THE PERSON AFFECTING RESTRICTION SOLVE THE PROBLEMS IN POPULATION ETHICS?

(Published in Roberts M. & Wasserman D. (eds.) *Harming Future Persons*, Ashgate, 2009.)

1. INTRODUCTION

The Person Affecting Restriction, in its slogan form, states that an outcome can only be better than another if it is better for someone.[1] It has a strong intuitive appeal and several theorists have suggested that the counterintuitive implications in population ethics of so-called "impersonal" welfarist theories, such as classical utilitarianism's implication of Derek Parfit's well-known Repugnant Conclusion, could be avoided by adopting the restriction.[2]

However, the Person Affecting Restriction has been subjected to some severe criticism by, among others, Derek Parfit, John Broome, Larry Temkin, and me.[3] This criticism turns around the implications of the restriction in non-identity cases, that is, cases in which the identity of the people that will exist in the future is at stake. It has been shown that in a number of such cases, the restriction has highly counterintuitive implications and yields non-transitive orderings.

Many have taken this criticism to be decisive and thus accept that we have to abandon the restriction. Recently, however, there have been some interesting reformulations and defences of the restriction, including one of my own making, suggesting that the restriction survives the "old" criticism. I shall investigate the viability of these proposals.

These new theories are all versions of a view I call Comparativism so I shall start by introducing this view. I shall then discuss different versions of the restriction and how these are related to Comparativism. The rest of the paper discusses the implications of different versions of Comparativism in population ethics. I shall show that most versions of this view either have counterintuitive implications or are extensionally equivalent with impersonal theories. The exception is a version of what I call Soft Comparativism. Albeit not without its own problems, it seems to have an advantage over impersonal theories.

2. COMPARATIVISM

According to Comparativism, we should draw a distinction between *uniquely* and *non-uniquely realisable* people. The former people only exist in only one out of two compared outcomes, whereas the latter exist in both of the compared outcomes.[4] The idea is that we should take the well-being of non-uniquely realisable people into account in a different way as compared to the well-being of uniquely realisable people.

Consider the following condition for same sized population:

> *Neutrality*: If there is a one-to-one correspondence from outcome A to outcome B such that every person in A has the same welfare as their counterpart in B, then A and B are equally good.

Standard welfarist axiologies, such as the axiological part of Total Utilitarianism, count everyone's welfare equally and thus satisfy Neutrality. A comparativist, however, counts people's welfare differently depending on whether they are uniquely or non-uniquely realisable and thus violates Neutrality.

A strict comparativist only counts the welfare of non-uniquely realisable people and completely disregards the welfare of uniquely realisable people. A soft comparativist, on

the other hand, counts the welfare of everybody but gives more importance to the welfare of non-uniquely realisable people.

According to a third view, among people with positive welfare, we should only count the welfare of non-uniquely realisable people, but among people with negative welfare, we should count the welfare of everybody. In other words, this view respects Neutrality in regard to populations with negative welfare. The reason behind this move is to incorporate an idea called *Asymmetry*: we have no moral reasons to create people with positive welfare, other things being equal, but we have reasons not to create people with negative welfare, other things being equal.

In many cases, the motivation behind drawing one or the other of the above distinctions is the Person Affecting Restriction. The idea is that the restriction is supposed to entail one or another of the above distinctions. How this entailment is supposed to work is by no means clear and depends, of course, on how one understands the Person Affecting Restriction, to which we shall now turn.

3. THE PERSON AFFECTING RESTRICTION

The Person Affecting Restriction has been formulated in different ways by different authors. I shall take as my point of departure the slogan above, originally formulated by Temkin. It is quite vague, however, and open to several interpretations, and I think it can be worthwhile to discuss some of them, since it might be that the intuitive support for the Person Affecting Restrictions derives from these weaker and less controversial principles.

The restriction could be understood as an idea about which kind of objects that moral goodness supervenes on, for example, that goodness is essentially related to the interests of human beings. All moral claims would thus necessarily involve a reference to humans: outcome A is better than outcome B since people have higher welfare in the former as

compared to the latter outcome, or since in the former but not in the latter outcome people's rights are fulfilled, or in the former but not in the latter people have equal opportunities, and so forth.

Examples of putative moral claims which are ruled out by this restriction would thus be: outcome A is better than outcome B since the scenery is beautiful in the former but ugly in the latter outcome, or since the ecosystem is in balance in the former but not in the latter outcome, and so forth. Roughly, this interpretation of the Person Affecting Restriction, which we could call the Human Good Restriction, claims that two outcomes can only differ in value if they differ in regard to some aspect of human goods.[5]

This restriction is pretty reasonable and I think that much of the appeal of the Person Affecting Restriction derives from the Human Good Restriction.[6] It is clearly insufficient, however, to yield any kind of distinction between the value of people that exist in more than one outcome and uniquely realisable people.

One can give a stronger interpretation of the Person Affecting Restriction than the one given above. One can stress an individualist aspect of value: all moral goods are *personal goods* which, roughly, are goods attached to individuals. These goods are features of people's lives that make them go better or worse. One version of this view is a form of Welfarism which claims that an outcome can only be better than another outcome if there is someone in the first outcome with a higher welfare level than someone in the other outcome (given the same number of people in the compared outcomes).

Consider the following two outcomes: in A, Krister and Erik are equally happy. In B, they are both happier than in A but Krister is happier than Erik. An egalitarian might argue that B is worse, or at least in one respect worse, than A, since although both Erik and Krister are better off in A than in B, B involves inequality whereas there is perfect equality in A. One might say that B is worse in regard to one aspect of human goods, namely its

distribution. “Worse for whom?” some theorists ask rhetorically. Perhaps they endorse a reading of the Person Affecting Restriction, which we could call the Personal Good Restriction, to the effect that an outcome cannot be worse than another, if it isn’t worse in regard to personal goods.[7]

The egalitarian concern above is grounded in a good that isn’t attached to one individual but to a relation between two individuals: What is bad about outcome B is that one person is worse off than another person. Consequently, this concern is ruled out by the Personal Good Restriction. Since B is not worse than A in respect to personal goods, B cannot be worse than A. In other words, if we find this restriction plausible, then we have another reason for rejecting Welfarist Egalitarianism.[8]

The Personal Good Restriction doesn’t imply, however, any value distinctions between uniquely and non-uniquely realisable people. It is compatible with such distinctions: one might decide, perhaps on purely intuitive grounds, that only personal goods belonging to non-uniquely realisable people count. It is, however, equally compatible with principles which don’t distinguish between uniquely and non-uniquely realisable people. Total Utilitarianism, for example, entails the Personal Good Restriction.

The next step to take is to stress the individualist aspect of value even more by claiming that morality is essentially *person comparative*: if an outcome is better (worse) than another, then it is better (worse) for at least one person. We shall formulate this view with a little bit more content.

The Person Affecting Restriction

(a) If outcome A is better (worse) than B, then A is better (worse) than B for at least one individual.[9]
(b) If outcome A is better (worse) than B for someone but worse (better) for no one, then A is better (worse) than B.[10]

This is the principle that I shall henceforth refer to as the Person Affecting Restriction (or "the restriction" for short). Strictly speaking, clause (a) above expresses what I take to be the restriction proper whereas clause (b) is a dominance condition. For expositional simplicity, I've added clause (b) to the statement of the restriction. It is quite an intuitively compelling condition in the present context and both the proponents and the opponents of the person affecting view in the literature seem to endorse such a dominance condition (I shall discuss a possible objection to it below).

It is not that the above statement of the restriction by itself captures what all authors have had in mind. As we shall see, however, most authors' views can be captured by the above restriction in combination with some further specification of what it means for an outcome to be better or worse for an individual.

In cases involving only the same people in the compared outcomes, this view is roughly extensionally equivalent to the Personal Good Restriction. In cases involving people whose existence is contingent on our choices, however, it becomes ambiguous. An outcome A is better than B for Peter if Peter has, for example, higher welfare in A as compared to B (we are assuming that if a person has higher welfare in one outcome as compared to another, then the former outcome is better for that person, other things being equal). But what if Peter exists in outcome A but not in outcome B? Is outcome A then better than outcome B *for Peter*? This is the crux of the matter. Depending on the answer to this question, different versions of Comparativism result.

4. STRICT COMPARATIVISM

One possible answer to the question whether existence can be better or worse for a person is to claim that non-existence is neither better nor worse than existence for a person: non-existence and existence are either equally as good or incomparable in value for a

person. This answer in combination with the restriction yields *Strict Comparativism*: we should completely disregard the welfare of *uniquely realisable* people, that is, people who only exist in one out of two compared outcomes.

This seems to be John Broome's take on the restriction. "Suppose [an alternative X] contains a certain number of people, and [an alternative Y] contains all the same people, and some more as well . . . . Then [the person-affecting view] is that [X] is at least as good as [Y] if and only if it is at least as good for the people who exist in both."[11]

At times, this also appears to be David Heyd's view. He argues that the welfare of future possible people has "no direct moral significance and cannot be decided in ethical terms."[12] Furthermore, he holds that "the very comparison of the welfare of two possible children is based on the fallacious notion of an abstract, impersonal quantity of happiness in the world which should be maximized" and that "[e]xcluding the welfare and interest of future merely possible person . . . is a necessary consequence of a coherent person-regarding theory of value." He suggests that we can solve the problems in population axiology "by simply rejecting the logical legitimacy of comparisons between the welfare of a possible population A and a possible population B (when they consist of *different* people)."[13]

Strict Comparativism has such counterintuitive conclusion that it is hard to believe that anyone seriously have endorsed it. Consider the *Bliss or Hell Case*:

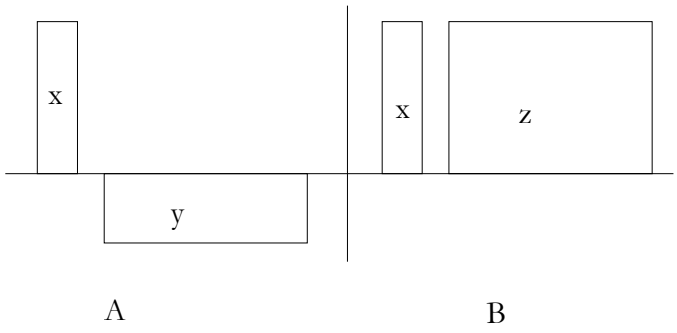

Diagram 1

Assume that we can either see to it that all the people in the future have hellish lives (the y-people in outcome A) or that they have excellent lives (the z-people in outcome B). Assume further that these two possible futures consist of different but the same number of people and that these two outcomes are equally good for us, the present x-people. Most of us, I presume, would consider outcome B clearly superior to outcome A and that we ought to realise B rather than A.

Since the y- and z-people are uniquely realisable people, outcome A and B are incomparable or equal in value for these people. Consequently, outcome A is neither better nor worse for the y- and z-people as compared to B. Moreover, the two outcomes are equally good for the x-people. Hence, according to the first clause of the Person Affecting Restriction, A cannot be better than B since it is not better for at least one individual (and vice versa). Consequently, Strict Comparativism ranks these outcomes as either equally

good or as incomparable in value. But that is clearly the wrong answer to the Bliss or Hell Case.

5. ASYMMETRICAL COMPARATIVISM

Strict Comparativism is a distraction. A perhaps more intuitively appealing answer to the question whether existence can be better or worse for a person is to claim that existence with positive welfare is equally as good as non-existence for a person, or incommensurable in value for a person, but a life with negative welfare is worse for a person than non-existence.

Melinda Roberts seems to have something like this in mind when she writes that "the choice to *bring* extra people into existence is . . . a morally neutral choice, so long as . . . the only people who are affected by that choice are the extra people themselves and those extra people themselves have *lives worth living"*[14] whereas "the bringing into existence of a person whose life is likely to be anguished is one that can indeed be coherently said to harm the victim, with the precise measure of harm consisting of the difference between the zero level of wellbeing implied by nonexistence and the negative level likely to be suffered."[15]

Asymmetrical Comparativism yields the right answer in the Bliss or Hell Case but it will fall foul for a version of it:

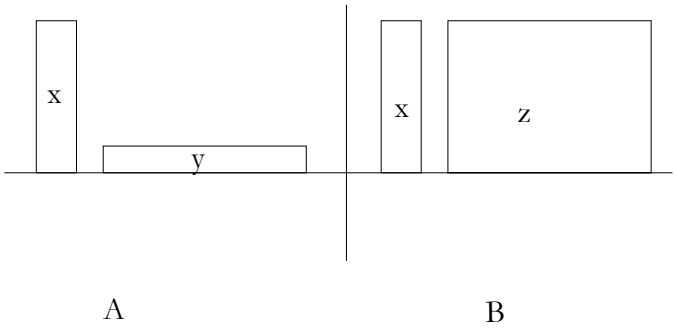

Diagram 2

In outcome A, the future y-people have lives barely worth living whereas the future z-people in B enjoy the same very high welfare as the x-people in B.[16] Again, since the y- and z-people are uniquely realisable people, outcome A is neither better nor worse for the y- and z-people as compared to B. As before, the two outcomes are equally good for the x-people. Hence, according to the first clause of the Person Affecting Restriction, A cannot be better than B since it is not better for at least one individual (and vice versa). Consequently, although A involves inequality and is much worse than B in regards to people's welfare, Asymmetrical Comparativism ranks these outcomes as either equally good or as incomparable in value. Again, we have reached the wrong answer.

In addition, it is not clear that it is possible to uphold this kind of asymmetry. The standard argument for why existence isn't better for a person than non-existence is well expressed by John Broome:

> [I]t cannot ever be *true* that it is better for a person that she lives than that she should never have lived at all. If it were better for a person that she lives than that she should never have lived at all, then if she had never lived at all, that would have been worse for her than if she had lived. But if she had never lived at all, there would have been no her for it to be worse for, so it could not have been worse for her.[17]

This argument works, however, equally well against the idea that existence could be worse for someone than non-existence. If it were worse for a person that she exists than that she should never have existed, then it would have been better for her if she had never existed. If she had never existed, then there would have been no her for it to be better for, so it could not have been better for her. Thus, it cannot be true that it could be worse for a person to exist than not to exist. In other words, it doesn't look possible to uphold an asymmetry here by an appeal to Broome's argument.[18]

6. INCONSISTENCY

Let's consider a last problem for Strict and Asymmetrical Comparativism which also will be of interest when we consider Soft Comparativism below. A common charge against the Person Affecting Restriction is that it yields non-transitive orderings. This is indeed true for Strict and Asymmetrical Comparativism, since taken as transitive population axiologies, they are inconsistent. Consider the following case:

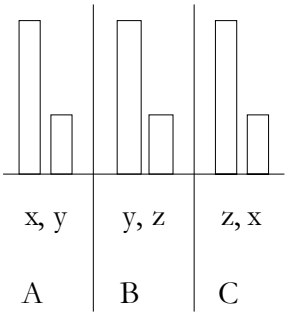

Diagram 3

The x- and y-people exist in outcome A, the y- and z-people exist in B, and the z- and x-people exist in C. Assume that all of these people have positive welfare, but that the

y-people are better off in B as compared to A, the z-people are better off in C as compared to B, and the x-people are better off in A as compared to C.[19]

Since the x-people don't exist in B, B is neither worse nor better than A for them. Similarly, since the z-people don't exist in A, A is neither worse nor better than B for them. However, B is better than A for the y-people. Consequently, B is better than A according to the second clause of the Person Affecting Restriction. The same reasoning yields that C is better than B, and A is better than C. But if B is better than A, and C is better than B, then transitivity yields that C is better than A. Consequently, A is both better and worse than C, which cannot be true.

The earlier arguments against Strict and Asymmetrical Comparativism relied only on clause (a) in the Person Affecting Restriction. The above argument, however, involved the dominance condition in clause (b). Thus, a strict or asymmetrical comparativist might here reply that the above inconsistency only follows because we have attributed a too strong dominance condition to her. Rather, she might claim, a strict comparativist believes that if outcome A is better than B for some people, and *at least as good* for everybody else, then A is better than B, but if A and B are incomparable in value for someone, then A and B are also incomparable in value.

This is an untenable answer for several reasons. Firstly, this theory yields strongly counterintuitive conclusions that also are at odds with the person affecting intuition. Consider the following case. In outcome D, there are a vast number of people, the x-people, suffering horribly. In E, all the x-people have excellent lives and there is an extra person, also with an excellent life. With the revised dominance clause above, Strict and Asymmetrical Comparativism implies that D and E are incomparable in value since it is incomparable in value for the extra person in E, and although all the non-uniquely realisable x-people are much better off in E as compared to D. This is hardly believable.

In response to this objection, the strict comparativist might reply by further revising the dominance clause and claim that if A and B are incomparable in value for some sufficient number of people, then A and B are also incomparable. But this position leads to more problems than it explains, it seems to me.

Firstly, even if there some great number of extra people with excellent lives in outcome E, it is still very counterintuitive, in my mind, that D and E are incomparable, and at odds with the person affecting intuition.

Secondly, this kind of comparativist owes us an explanation how it can be that the incomparability of outcomes is a function of the number of people for which the outcomes are incomparable in value. To the best of my knowledge, there is no such credible account.

Finally, this theory is a strange hybrid which I think falls between the chairs of Strict and Soft Comparativism. The motivating idea behind Strict Comparativism is that since the relative goodness of outcomes are determined only by how they are better or worse for individuals, and since outcomes cannot be better or worse for uniquely realisable people, we can ignore them when ranking outcomes. According to the hybrid theory, however, we cannot ignore uniquely realisable people, since sometimes their mere existence can make outcomes incomparable. On the other hand, and in contrast with the soft comparativist, we can ignore the *welfare* of uniquely realisable people since it is irrelevant for the goodness of outcomes.

7. PERSONAL GOOD RESTRICTION REGAINED

The last answer to the question whether existence can be better or worse for a person that I shall consider is the claim that existence with positive welfare is better for a person than non-existence and that existence with negative welfare is worse for a person than non-existence.

As my discussion above indicates, I worry that such statements really don't make sense when the "better for" relation is understood in ordinary welfarist terms, that is, in terms of what is better or worse for the individual concerned rather than in terms of what we ought to do or what makes the world better or worse.

There is, however, a way of analyzing the "better for" relation that makes sense of this idea although it might not capture all aspects of the ordinary notion of "better for."[20] One could explicate "better for" in terms of what a benevolent impartial observer or guardian angel would choose for a person when she is only considering the wellbeing of the person under consideration (as all guardian angels do). According to this view, a state X as better for a person than state Y if and only if this is what a guardian angel would choose for her. If the guardian angel has a choice between bringing someone into existence with negative welfare or not bringing her into existence at all, she would of course choose the latter. Moreover, it doesn't seem unreasonable to say that if the guardian angel had the choice between bringing someone into existence with positive welfare and not bringing her into existence, she would choose the former. On this view, this is all that it means to say that non-existence is better or worse for a person.[21]

Consequently, given these understandings of the "better for" relation, one can claim that it is worse for a person to exist with negative welfare than not to exist at all, and that it is better for a person to exist with positive welfare than not to exist at all, without implying any absurdities.

In combination with the restriction, this answer to the question whether existence can be better or worse for a person doesn't have any of the disagreeable implications of the versions discussed above. It doesn't have any force, however, and doesn't imply any value distinctions between uniquely and non-uniquely realisable people. Actually, it is just a

restatement of the Personal Good Restriction discussed earlier.[22] Hence, it is compatible with, for example, Total Utilitarianism and other axiologies that respect Neutrality.

On a more positive note, the reasoning above shows how one can avoid the pejorative rhetoric of an "impersonal ethics," which is often applied to Total Utilitarianism and other views that respect Neutrality, since it is possible to couch one's principles in person affecting terms but still count everybody's welfare equally.

8. A NORMATIVE VERSION OF THE RESTRICTION

A possible rejoinder to the non-transitivity problem for Comparativism presented in Diagram 3 is to argue that the restriction should be couched in normative rather than axiological terms, supplemented with the plausible claim that there is no analogue to the transitivity of "better than" for those normative concepts.[23]

Here's how we shall formulate a normative version of the restriction:

The Normative Person Affecting Restriction:

(a) If it is impermissible to choose outcome A in a certain situation, then there is another outcome B in the situation such that A is worse than B for at least one individual.

(b) If outcome A is better than B for someone but worse for no one, then in any situation involving a choice between A and B, it is wrong to choose B.

As with the axiological version of the restriction, clause (a) above expresses what I take to be the normative restriction proper whereas clause (b) is a weak dominance condition.

We can now combine this restriction with different answers to the question whether existence can be better or worse for a person to get different normative versions of

Comparativism. Assume now that we combine it with the claim that non-existence and existence are equal or incomparable in value for a person to get a normative version of Strict Comparativism. As is easily seen, this view implies that none of the outcomes in the Bliss or Hell Case are impermissible because of clause (a) in the Normative Person Affecting Restriction. What will be the implication of this theory in regard to the case in Diagram 3?

Again, since the x-people don't exist in B, B is neither worse nor better than A for them. Similarly, since the z-people don't exist in A, A is neither worse nor better than B for them. However, B is better than A for the y-people. Consequently, A is an impermissible choice according to the second clause of the Normative Person Affecting Restriction. The same reasoning applied to outcome B and C yields that these outcomes are also impermissible choices. Hence, Normative Strict Comparativism implies that all the available alternatives are wrong and we are facing a moral dilemma: whatever act we perform we are going to act wrongly.[24]

I'm sceptical about such moral dilemmas in general, but be that as it may, I think it is unlikely that the case described in Diagram 3 can credibly be described as a moral dilemma. It would be odd to claim that, if faced with the situation in Diagram 3, we would do something wrong if we choose, say, outcome A. The claim is odder still if we assume that everyone in Diagram 3 enjoys very high welfare in all outcomes and that the difference between the best and worst off is small (notice that this assumption wouldn't change the structure of the case).

Moreover, we would get a moral dilemma even if the y-people are only slightly worse off in A, whereas the z- and x-people are much worse off in B and C, respectively. In such a case, it seems clear that it is at least permissible to choose A since it involves

much less harm and much more well-being than B and C. Yet, according to Strict Normative Comparativism, all the alternatives are wrong.

Lastly, consider a version of the case in Diagram 3 where there is another outcome F consisting only of people with very low positive welfare, the w-people. Again, Normative Strict Comparativism implies that alternatives A, B and C are wrong. In addition, the first clause of the restriction implies that it cannot be impermissible to choose outcome F, since there is no alternative outcome in which the w-people are worse off. Hence, assuming that there are no deontic gaps, Normative Strict Comparativism implies that alternative F is obligatory.

The above discussion points to a general problem for those who hope to evade non-transitive orderings by switching to a normative framework. As I have showed elsewhere, non-transitive value orderings most often translate into moral dilemmas on the normative level.[25]

Melinda Roberts seems to apply this strategy when she tries to show that in cases where the axiological version of the restriction yields non-transitive orderings (which she calls the "inconsistency argument" against the Persons Affecting Restriction), a normative interpretation of the restriction will yield reasonable results.[26] She shows that this can be done in some versions of the Mere Addition Paradox. To that end, she formulates a theory called *Personalism* ("W(p, X)" is the numerical representation of person p's welfare in outcome X):

> *Personalism*:
>
> P1: An alternative X is impermissible, or wrong, for agents if and only if some existing or future person is or will be wronged by agents at X.
>
> P2: A person p is not wronged at X if, for each Y that is an alternative to X, $W(p, X) \geq W(p, Y)$.

P3: A person *p* is not and will not be wronged at X if *p* does not and will not exist at X.

P4: A person *p* is wronged at X if
- (i) there is some Y such that W(*p*, Y) > W(*p*, X),
- (ii) there is no *q* who exists or will exist at Y such that W(*q*, X) > W(*q*, Y) and
- (iii) for each *q* who exists or will exist at Y, *q* exists at some time at X.[27]

Roberts adds an implicit fifth principle to these four, namely that "[i]f no one is or will be wronged at a given alternative, then we know that the alternative itself represents a permissible choice."[28]

What implications does Personalism have for the inconsistency argument described in Diagram 3? Consider the y-people in A. Clause (i) and (ii) in P4 are satisfied since the y-people are better off in B and there is no person in B who has higher welfare in A. However, clause (iii) is not satisfied since the z-people only exist in B. Consequently, because of clause (iii), Personalism is silent about this case.

This result makes Roberts' solution to the inconsistency argument seem a bit unsatisfactory. A satisfactory solution to the inconsistency argument should show, I think, that in all cases where the axiological restriction yields non-transitive orderings, the normative version yields a somewhat reasonable evaluation. Moreover, the case described in Diagram 3 is the classical argument for the non-transitivity of the Person Affecting Restriction.

Notice also that Personalism is silent in cases where it seems clear what we should and shouldn't choose from a person affecting viewpoint. Assume that, in another situation, we have a choice between only outcome D and E:

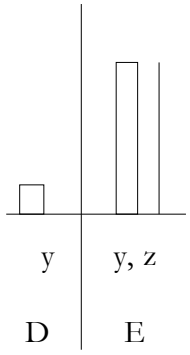

Diagram 4

Outcome D consists of the y-people with very low positive welfare whereas outcome E consists of the y-people and one z-person, all with very high welfare. Hence, the y-people are much better off in E as compared to D and, intuitively, the y-people would be very much wronged if we were to choose outcome D. Moreover, there is an extra person enjoying very high positive welfare in E. Clearly, the right result here is that D is impermissible and E is permissible. Personalism, however, neither implies that the y-people are wronged nor that D is impermissible. Again, clause (iii) is not satisfied, this time because the z-person exists in only one of the compared outcomes.

One might think that Personalism would at least imply that E is permissible. As it is stated, however, it doesn't. To know that E is permissible, we must know that no one is wronged in E. It follows from P2 that the y-people are not wronged in E. But it doesn't follow from P2 or any of the other principles that the z-people are not wronged in E. It doesn't follow from P2 since it is not true for any z-person z_i that $W(z_i, E) \geq W(z_i, D)$ because the z-people doesn't have a welfare level in D as they don't exist there.

In Roberts' discussion of other examples, it appears that she does assume something to the effect that one is not wronged by being brought into existence with positive welfare as long as the alternative is non-existence or existence with lower welfare.[29] One way to

achieve this result is to claim, a bit nonsensically in my mind, that one has zero welfare in the outcomes in which one doesn't exist.[30] In another paper, Roberts' seems to claim exactly this: "Nora does not have any properties at all at any alternative at which she does not exist, and, second, that, where Nora has no properties at all, all the properties she does have—the empty set—add up to a zero level of wellbeing."[31] With this assumption, P2 will yield the desired result and E will be permissible. Still, however, it doesn't follow that the y-people are wronged in D and this outcome is not classified as impermissible.

Since Roberts has only stated a sufficient condition for when people are wronged, she could respond by adding some condition to the effect that the y-people are wronged in D in the above diagram and in A in Diagram 3. Such a condition would also imply that the z-people are wronged in B in Diagram 3, and that the x-people are wronged in C (we would get the same result if we dropped clause (iii) in P4). It follows from P1 that all the available alternatives are wrong and we are again facing a moral dilemma.

Such a version of Personalism would also imply moral dilemmas in all the cases above in which Normative Strict Comparativism implied such dilemmas. Moreover, it would imply that outcome F is obligatory in the last case that we discussed in connection with Normative Strict Comparativism.

A better response is to add some condition to the effect that the y-people are not wronged in A in Diagram 3, and likewise for the z-people in B and the x-people in C. We would then get the result that outcome A, B, and C, are all permissible, which seems intuitively correct.[32]

A rather intuitive way of achieving this result, I think, is to claim that the y-people are harmed in A although this is not sufficient to settle the matter whether the y-people are wronged or not and whether A is permissible or not, since other people might be harmed (e.g., the z-people in B) in the alternative outcomes. To settle which alternatives are

permissible, we would then have to weigh together how much people are harmed in the different outcomes. Moreover, since it seems reasonable to claim that the y-people are as much harmed in A as the z-people in B and the x-people in C, such a view could yield that all the outcomes have the same deontic status and that they are all permissible. This view is similar to what I call Soft Comparativism and the Person Affecting Principle of Comparative Harm which we shall discuss in the next section.

9. SOFT COMPARATIVISM

According to the soft comparativist we should count the welfare of everybody but give more weight to the welfare of non-uniquely realisable people in some manner. Here's how I think we can develop a kind of person affecting version of this theory. [33] The central idea is that a person is harmed if she is worse off than she could have been, and it makes an outcome worse if people are harmed, or it might make it impermissible to choose the outcome. Another way to put this idea is to say that such people have a legitimate complaint or grievance and this makes the outcome worse or might make it impermissible. In addition to the wellbeing of everybody, this proposal takes the badness of legitimate complaints, or comparative harms, into account.

Consider the following case from Parfit:

> *Two Medical Programmes*: If a pregnant woman has medical condition J, which a simple treatment could cure, this will cause the child she is carrying to have a certain handicap. If a woman has condition K when she conceives a child, this child would get the same handicap. Condition K cannot be treated but disappears after two months. There are two medical programmes: Pregnancy Testing for J (PTJ) and Preconception testing for K (PCTK). In PTJ, women would be tested during pregnancy and those found to have condition J would be treated. It is predicted that if we implement PTJ, 1000 children who would otherwise have been handicapped will be born without the handicap. In PCTK, women would be tested when they intend to become pregnant, and those found to have K would be advised to postpone conception for at least two months. It is predicted that if we implement PCTK, 1000 children will be born without the handicap rather than 1000 (different) handicapped children. We

> only have funds for one of the medical programmes. Which one should we choose?[34]

Since both programs would reduce the number of handicapped children by a 1000, many would, like Parfit, consider these programs equally good. Likewise, this is the implication of Total Utilitarianism and other theories that respect Neutrality.

If we choose to implement PCTK, however, there will be 1000 children with a handicap that they wouldn't have had if we had chosen to implement PTJ instead—these people are non-uniquely realisable and will exist irrespective of our choice. These people can therefore be said to have been harmed and thus have a legitimate complaint. If we choose to implement PTJ, there will be 1,000 children with a mild handicap but since these children owe their existence to our choice—they are all uniquely realisable people—they cannot be said to have been harmed or made worse off and thus don't have a legitimate complaint.

Consequently, although the effect on people's welfare is the same for both programs, PCTK is worse in one respect since it will cause people to be worse off than they could have been and thus there will be people who are harmed or can legitimately complain. Soft Comparativism captures this intuition and ranks PTJ as better than PCTK.

As is easily seen from our discussion above, Soft Comparativism doesn't follow from the Person Affecting Restriction as we have understood it here (it only follows that either only non-uniquely realisable people, or only non-uniquely realisable people and people with negative lifetime welfare count, or everybody counts and counts equally). I suggest that the interpretation of the person affecting idea that figures here is not a necessary condition for an outcome being better or worse but rather a condition for when people are harmed:

> *The Principle of Comparative Harm*: If a person would be worse off in an outcome A as compared to an alternative outcome B, then she would be comparatively harmed if A rather than B came about.[35]

So the person affecting version of Soft Comparativism is then the view that the value of an outcome, or the deontic status of the act yielding the outcome, is determined by both the value of the welfare in the outcome and the value of the comparative harm in the outcome. Henceforth, I shall just refer to this view as "Soft Comparativism."

As with the Person Affecting Restriction, we will get different versions of the comparative harm principle depending on which answer we give to the question whether existence can be better or worse for a person. If we claim that non-existence is neither better nor worse than existence for a person, then we get the result that people are not harmed by being brought into existence. Alternatively, we could again introduce an asymmetry and claim that existence with negative lifetime welfare is worse than non-existence for a person whereas existence with positive welfare is not better than non-existence for a person.

This is an interesting idea but there are problems ahead. Consider again Diagram 3. All the outcomes in the diagram are equally good in respect to the amounts of people's well-being. However, since the y-people are worse off in A as compared to B, the y-people would be harmed if A came about. In this respect, A is worse than B. Consequently, all things considered, A is worse than B. The same reasoning yields that B is worse than C, and C is worse than A. But if A is worse than B, and B is worse than C, then transitivity yields that A is worse than C. Again, it follows that A is both better and worse than C, which cannot be true.

There is, however, another way of explicating Soft Comparativism which doesn't imply non-transitive orderings. When determining the value of an outcome we should

consider both people's well-being and whether they are harmed in the sense of being worse off than they could have been. The value of an outcome is determined by the value of the well-being in the outcome reduced by a factor that reflects whether people are harmed.[36]

Here's an example of how this could be done. Assume that we represent well-being on a numerical scale and that the total wellbeing of the best-off people in Diagram 3 is 10 units and the total wellbeing of the worst-off people is 5 units. The value of outcome A would then be 15 minus some factor *h* that represent the fact that the y-people are worse off than they could have been. Intuitively, this factor should correspond to how much worse off the y-people are in A as compared to B.[37] Similarly, the value of outcome B and C would be 15 minus *h*. Consequently, on this view all the outcomes in Diagram 3 are ranked as equally good which seems to be the intuitively correct answer.

Since this version of Soft Comparativism assigns a specific numerical value to all outcomes, it cannot imply non-transitive orderings. Hence, we here have a kind of person affecting theory that refutes the inconsistency argument.

Moreover, in regard to the two medical programmes, Soft Comparativism would pick PTJ since the two programs are equally good in regard to people's welfare but PCTK is worse in one respect since it will cause people to be worse off than they otherwise could have been.

Lastly, in regard to the Bliss or Hell case in Diagram 1, irrespective of whether we consider the y-people harmed or not in outcome A, Soft Comparativism will rank B as better than A since it also takes into account the well-being of everybody, which is clearly superior in outcome B.

10. DOMINATED OUTCOMES

Although Soft Comparativism neatly captures some people' intuitions regarding the two medical programs and avoids the threat of non-transitivity, it also has implications which one might consider counterintuitive. Consider the following three outcomes:

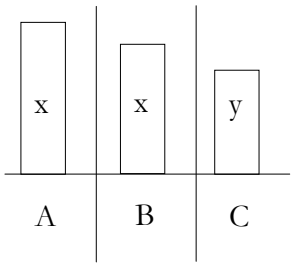

Diagram 5

Each of the three outcomes contains the same number of people. Everyone in A is better off than everyone in B, and everyone in B is better off than everyone in C. Again, the x-people would be harmed if outcome B came about since they would be worse off than they otherwise could have been, that is, if A had been the case instead of B. Let *h* represent the total value of the harm done to the x-people in B. Let *d* represent the total difference in well-being between the x-people in B and the y-people in C. The difference in value between outcome B and C will then be *d* minus *h*.[38] Consequently, if *h* is greater than *d*, then Comparativism will rank C as better than B although everyone in the latter outcome is better off than everyone in the former. This seems a bit counterintuitive.

Yet, this result might not perturb the comparativist. She might say that since she believes in the negative value of comparative harm, she's willing to trade off some welfare to avoid such comparative harm. She might point to other non-welfarist axiologies such as desert theories where the value of an outcome is determined both by the receipt of welfare and the fit between receipt and desert. On such theories, there might arise some cases

where we have to forego some welfare to achieve a better fit between receipt and desert.[39] Likewise for other pluralist axiologies.

One might think that this shows that Soft Comparativism really isn't a person affecting theory as we have understood it here. Notice again that everyone in C is worse off as compared to everyone in B. According to the Person Affecting Restriction, if outcome C is better than B, then C should be better than B for at least one individual, which it isn't. So, rather than being a person affecting theory, Soft Comparativism is a pluralist axiology that sometimes will trade off people's well-being against the value of avoiding comparative harms.

As we already noted above, however, Soft Comparativism doesn't follow from the Person Affecting Restriction. Again, Soft Comparativism is based on an interpretation of the person affecting view as an idea regarding when people are harmed in different outcomes.

There are, however, further problems with this view if we turn to its normative implications. Assume some weak form of consequentialism, or consequentialist part of a deontological theory, to the effect that we ought to choose the best outcome in the cases currently under discussion.[40] Assume further that A is a very unlikely outcome. If we try to achieve A, we are most likely to fail and end up with C. If we aim at B or C, we will succeed. Still, Soft Comparativism tells us that it would be wrong to choose B, just because there is an unlikely outcome A in the choice set.[41]

One might think that this problem can be fixed by letting the harm-factor depend on the probability of the better alternative. Instead of letting *h* represent the total value of the harm done to the x-people if we were to choose outcome B, it should be represented by *ph,* where *p* is the probability that A will be the case given that we choose A, and *h* is the harm done to the x-people were we to choose outcome B when we could have chosen A with

certainty of success. The difference in value between outcome B and C will now be *d* minus *ph* so still, if *ph* is greater than *d*, which is clearly possible, then Comparativism will rank C as better than B although everyone is better off in B.

A better solution might be to let the harm-factor depend on people's expected welfare given a certain action which with certain probabilities brings about certain outcomes. On this view, people are harmed if their expected welfare is lower than it could have been given a different choice of action. Assume that if we choose an action a_A aiming at bringing about A, then the probability that A will be the case is 0.10 and the probability that C will be the case is 0.90, whereas if we choose action a_B, then B will be the case with certainty. Assume further that the total well-being in A is ten and in B eight. Then the x-people's expected welfare if we choose a_A is 0.10 x 10 = 1, whereas it is 1 x 8 were we to choose action a_B. Consequently, on this formulation of Soft Comparativism, the x-people are not harmed if we choose a_B since their expected welfare is higher if we choose that action rather than a_A. Hence, by switching to expected welfare and defining harm in terms of expected welfare, the problem of unlikely outcomes disappears.

On the other hand, if we went for a_A and A actually came about, then we would still have harmed the x-people and done the wrong thing since their expected welfare (at the time of the choice) was lower than it would have been had we chosen a_B, although the x-people are better off since A rather than B actually came about. This might strike some as implausible. But a possible rejoinder is to claim that it was wrong to choose a_A since that action exposed the x-people to a risk of getting nothing.[42]

This problem, however, is not peculiar to Soft Comparativism but analogous to the old dispute among consequentialists regarding whether one should adopt a formulation of consequentialism in terms of the actual or probable outcomes of actions, so I shall not

discuss it further here.[43] It is noteworthy, however, that Soft Comparativism seems more compatible with a probabilistic rather than an actualistic formulation of consequentialism.

Let me end this section with two other objections to Soft Comparativism.[44] In Diagram 3, this theory correctly ranked all the outcomes as equally good. One might object, however, that we cannot know this without knowing that exactly these three outcomes are the only ones available in the situation since, according to Comparativism, the value of an outcome depends on the set of possible outcomes in the situation. Suppose, for example, that there was another outcome D with only the x-people at level 15. This would not only yield that D was the best outcome in the situation but also change the ranking of A, B, and C, since the x-people in C will be more harmed than the y-people in A and the z-people in B. Hence, C will be ranked as worse than A and B.

The first objection is that it is absurd that one and the same outcome can both be worse than and equally as good as another outcome. This seems to be the case here since, when D is not present in the set of outcomes, C is ranked as equally as good as A and B whereas, when D is present, C is ranked as worse than A and B. Hence, it looks like the same outcome, C, is both worse than and equally as good as A and B.

This would surely be absurd but the obvious rejoinder is to deny that these outcomes are the same outcomes. We can just partly individuate outcomes by the situation to which they belong. Hence, if we add another outcome to the situation described in Diagram 3, then we have a new situations with, say, alternatives A', B', C' and D and it is B' which is better than C' which doesn't contradict that outcome B and C in the original situation are equally good.

The second objection is that in practice we could never be epistemically justified in limiting the number of possible outcomes as we have done in the examples above. Hence,

since the Soft Comparativist ranking depends on the possible outcomes in the situation, we cannot be justified in believing in the ranking.

It is true that this makes Soft Comparativism a bit special as an axiology since most axiologies, such as the axiological component of classical utilitarianism, yield context-insensitive rankings of outcomes. However, this problem appears for these theories on the normative level, since which outcome is the best one, and thus which one we ought to choose, depends on which other outcomes are available in the situation. Hence, this alleged particular problem with Soft Comparativism reduces to the old problem of whether consequentialist theories ought to be and can be action-guiding and is thus no special problem for Soft Comparativism. The same standard responses come in handy here. For example, we could make a sharp distinction between criterion theories and decision methods and claim that Soft Comparativism is a criterion theory that has no claim to be used as a decision method other than indirectly in the choice of which decision methods that we should use in everyday life.[45]

11. FUTURE POPULATIONS AND TRADE OFFS

So far, so good. The main problem with Soft Comparativism is, however, that it will not be of much help in regard to other problems in population axiology. A difficulty shared by all versions of Soft Comparativism is that all the problems afflicting neutral theories will reappear in the specification of the method for aggregating people's welfare. Summing implies Parfit's Repugnant Conclusion; averaging implies that it can be better to create miserable rather than happy people, and so forth.[46]

For instance, assume, as we did above, that the value of an outcome is determined by the value of the total well-being in the outcome reduced by a factor that reflects whether people are harmed. In all cases involving only uniquely realisable people, Soft

Comparativism determines the ranking by the total sum of people's welfare since such cases don't involve any harm. Consequently, like Total Utilitarianism, it will imply the Repugnant Conclusion in respect to future populations where there is no overlap of individuals in the compared populations.

This might be an acceptable implication for some comparativists, however, since what matters to them is that people are not worse off than they otherwise could have been. Nevertheless, Soft Comparativism will imply similar conclusions even in cases that involve overlaps and thus involve great losses for non-uniquely realisable people (e.g., present people) and a lot of comparative harm.

Here's a numerical illustration of this point, involving a micro-version of the Repugnant Conclusion:

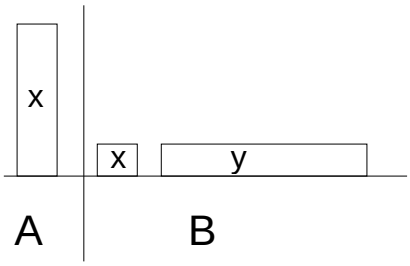

Diagram 6

Assume that we have a choice between outcome A with ten persons, the x-people, and outcome B with the x-people and an additional two hundred persons, the y-people. In outcome A, the x-people have very high lifetime welfare. Assume that this high welfare corresponds to a ten units of welfare. Consequently, the value of outcome A is 10 x 10 = 100.

In outcome B, the x- and y-people have very low positive lifetime welfare. Assume that this very low welfare corresponds to one unit of welfare. Since the x-people have much lower welfare in B, they are harmed in B. Assume that the harm factor for each x-person corresponds to her difference in welfare between outcome A and B. Thus, the value of the harm to the x-people in B is 10 x (-9) = -90 whereas the value of their welfare is 10 x 1 = 10. Taken all together, the value of the x-people in outcome B is 10 – 90 = -80.

However, since there are also two hundred y-persons in outcome B, the total value of outcome B, according to Soft Comparativism, is 200-80 = 120 which is greater than the value of A. Hence, Soft Comparativism here opts for the outcome with very low positive welfare, outcome B, although it involves a great loss in welfare and a lot of comparative harm for the x-people, and thus implies the Repugnant Conclusion. In this respect, Soft Comparativism doesn't constitute much of an advance towards a satisfactory population axiology as compared to standard welfarist axiologies such as Total Utilitarianism.

One might think that this result depends on the weight given to comparative harm. However, as long as the harm factor is represented by a finite number (that is, as long as we don't give lexical priority to comparative harm), Soft Comparativism will have the above implication. Here's a general demonstration:

Assume that h is a positive finite number that represents the weight given to the comparative harm of an individual due to the fact that she is worse off than she could have been. Let A consist of n non-uniquely realisable persons with very high welfare u_1. Let B consist of a mixed population of n non-uniquely and m uniquely realisable people with very low positive welfare u_2. The value of A is thus nu_1 and the value of B is $n(u_2 - h) + mu_2$. Now, for any value of h, there is an m such that $nu_1 < n(u_2 - h) + mu_2$, that is, a value of m that makes B better than A, namely $m > n(u_1 - u_2 + h)/u_2$.

If we give lexical priority to avoiding comparative harm, then the Repugnant Conclusion will be avoided. Given such an extreme negative weight on comparative harms, however, Soft Comparativism will yield counterintuitive result in a version of the case described in Diagram 2, and will face an extreme version of the case described in Diagram 5.

Recall that in Diagram 2 the y-people in A have lives barely worth living and the z-people in B enjoy the same very high welfare as the x-people in A and B. Assume further that one of the x-persons has slightly lower welfare in B as compared to A. If one gives lexical priority to avoiding comparative harms, then A is better than B according to Soft Comparativism, since no one is harmed in A but one of the x-people is harmed in B.[47]

Likewise, consider a version of the case depicted in Diagram 5 in which the y-people in C have lives barely worth living and only one of the x-people have slightly lower welfare in outcome B as compared to outcome A. Again, if we give lexical priority to avoiding comparative harms, C is better than B.

Soft Comparativism might be caught in a dilemma here. Only if it gives lexical negative weight to comparative harm will it avoid the Repugnant Conclusions in cases that involve great losses for non-uniquely realisable people and thus a lot of comparative harm. However, with such a negative weight on comparative harms, it will imply clearly unacceptable versions of the cases described in diagrams 2 and 5. It seems hard to find an acceptable way out of this dilemma.[48]

11. SUMMARY

The Person Affecting Restriction has a strong intuitive appeal. It is often contrasted with Total Utilitarianism and other "impersonal" theories and it has been hoped that the

restriction would help us avoid the disagreeable implications of impersonal theories in population ethics. I've argued that there are some versions of the idea that are quite acceptable. On the other hand, these versions are compatible with Total Utilitarianism and other neutral theories that count everybody's welfare and count it equally. Hence, these theories cannot help us solve the problems in population ethics but, on the other hand, these versions can help us to escape the pejorative rhetoric of an "impersonal ethics" since they can be used to formulate our principles in person affecting terms but still count everybody's welfare equally.

The main part of the paper considered Comparativism, the view that we should take the well-being of non-uniquely realisable people into account in a different way as compared to the well-being of uniquely realisable people. As I showed, different versions of Comparativism can be derived from one version of the Person Affecting Restriction, depending on how we answer the question whether existence can be better or worse for a person.

With one exception, all of these theories have very counterintuitive implications. The exception was a combination of the Person Affecting Restriction with the claim that existence with positive welfare is better for a person than non-existence and that existence with negative welfare is worse for a person than non-existence. Again, we obtained a theory which is compatible with Total Utilitarianism and other neutral theories.

Lastly, I considered Soft Comparativism according to which we should count the welfare of everybody but give more weight to the welfare of non-uniquely realisable people in some manner. Soft Comparativism cannot be derived from the Person Affecting Restriction but, as I suggested, it can be partly based on an interpretation of the person affecting idea as a condition for when people are harmed, namely the Principle of Comparative Harm. According to that principle, if a person would be worse off in an

outcome A as compared to an alternative outcome B, then she would be comparatively harmed if A rather than B came about.

As we showed, Soft Comparativism has a number of attractive properties. It avoids the inconsistency charge that has been levelled against the Person Affecting Restriction. Moreover, it yields the right answer in a number of test cases, including cases where Total Utilitarianism seems to give the wrong answers, such as the Two Medical Programmes. We also showed that a number of possible arguments against Soft Comparativism could be rebutted. Nevertheless, as we showed in the last section, Soft Comparativism runs into problems with the Repugnant Conclusion and cases that involves tradeoffs between the welfare of non-uniquely and uniquely realisable people.

Department of Philosophy, Stockholm University
Swedish Collegium for Advanced Study
Gustaf.Arrhenius@philosophy.su.se

References

Arrhenius G. 2000a. *Future generations: A challenge for moral theory*. Uppsala: University Printers (F.D. Diss.).

—. 2000b. An impossibility theorem for welfarist axiologies. *Economics & Philosophy* 16*:* 247-266.

—. 2003. Feldman's desert-adjusted utilitarianism and population ethics. *Utilitas* 15(2): 225-236.

—. 2003b. The person affecting restriction, comparativism, and the moral status of potential people. *Ethical Perspectives* 10 (3-4): 185-195.

—. 2005. The paradoxes of future generations and normative theory. In *The repugnant conclusion* (eds.) J. Ryberg and T. Tännsjö. Dordrecht: Kluwer Academic Publisher: 201-18.

—. 2006. Desert as fit: An axiomatic analysis. In Richard Feldman, Kris McDaniel, Jason R. Raibley and Michael J. Zimmerman, eds. *The good, the right, life and death: Essays in honor of Fred Feldman*. Aldershot: Ashgate Publishing.

—. 2006b. The moral status of potential people," Dept. of Philosophy, Stockholm University (http://people.su.se/~guarr/) (unpublished manuscript).

—. 2006c (together with Jesper Ryberg and Torbjörn Tännsjö) "The Repugnant Conclusion", in Edward N. Zalta (ed.) *The Stanford Encyclopaedia of Philosophy*, <http://plato.stanford.edu/archives/spr2006/entries/repugnant-conclusion/>.

—. 2007. Meritarian axiologies and distributive justice. In *Hommage à Wlodek: Philosophical papers dedicated to Wlodek Rabinowicz* (eds.) T. Rønnow-Rasmussen, B. Petersson, J. Josefsson and D. Egonsson, 1-43, www.fil.lu.se/hommageawlodek.
—. 2008. Life extension versus replacement. *Journal of Applied Philosophy* 25(3): 211-227.
—. 2009. Egalitarianism and population change. In *Theories of intergenerational justice* (eds.) A. Gosseries & L. Meyer. Oxford: Oxford University Press, forthcoming.
Bales, R. E. 1971 "Act-Utilitarianism: Account of Right-Making Characteristic or Decision-Making Procedure?", *American Philosophical Quarterly*, 8(3): 257-265.
Brink, D. 1986 "Utilitarian Morality and the Personal Point of View", *Journal of Philosophy*, 83(8): 417-38.
Broome, J. 1991. *Weighing goods: Equality, uncertainty, and time*. Oxford: Basil Blackwell.
—. 1992. *Counting the costs of global warming*. Cambridge: White Horse.
—. 1999. *Ethics out of economics*. Cambridge: Cambridge University Press.
—. 2004. *Weighing Lives*. Oxford: Oxford University Press.
Bykvist, K. 1998. *Changing preferences: A study in preferentialism*. Uppsala University (F.D. Diss.).
—. 2007. The Benefits of Coming into Existence. *Philosophical Studies*, 135:335–362.
Carlson, E. 1995. *Consequentialism Reconsidered*, Dordrecht: Kluwer Academic Publisher.
Danielsson, S. 1974. *Two Papers on Rationality and Group Preferences*, Filosofiska Studier no. 21, Uppsala: Filosofiska föreningen och filosofiska institutionen.
Feldman, F. 1997. *Utilitarianism, Hedonism, and Desert: Essays in Moral Philosophy*. Cambridge: Cambridge University Press.
——— 2006. "Actual Utility, the Objection from Impracticality, and the Move to Expected Utility", *Philosophical Studies*, 129:49–79.
Glover, J. 1977. *Causing death and saving lives*. Harmondsworth: Penguin Books.
Heyd, D. 1988. Procreation and value: Can ethics deal with futurity problems. *Philosophia* (Israel) 18: 151-170.
—. 1992. *Genethics: Moral issues in the creation of people*. Berkeley: University of California Press.
Holtug, N. 1996. In defence of the slogan. In *Preference and value: Preferentialism in ethics* (ed.) W. Rabinowicz, 64-89, Lund: Lund University (Dept. of Philosophy).
Hope, T. 2003. Physicians' duties and the non-identity problem. Oxford University (unpublished manuscript).
Kymlicka, W. 1990 *Contemporary Political Philosophy*, Oxford: Clarendon Press.
Meyer, L. 2003. Intergenerational Justice. *Stanford encyclopaedia of philosophy*.
Moore, G. E. 1903 [1966]. *Principia ethica*. Cambridge: Cambridge University Press.
Narveson, J. 1967. Utilitarianism and new generations." *Mind* 76: 62-72.
Parfit, D. 1984 [1991]. *Reasons and persons*. Oxford: Clarendon Press.
Rabinowicz, W. "Broome and the intuition of neutrality", *Philosophical Issues*, forthcoming.
Roberts, M. A. 1998. *Child versus childmaker: Future persons and present duties in ethics and the law*. Lanham, Maryland: Rowman and Littlefield.
—. 2003a. Can it ever have been better never to have existed at all? Person-based consequentialism and a new repugnant conclusion. *Journal of Applied Philosophy* 20(2): 159-185.

—. 2003b. Is the person-affecting intuition paradoxical? *Theory and Decision* 55: 1-44.
—. 2007. The non-identity fallacy: Harm, probability and another look at Parfit's depletion example. *Utilitas* 19(3): 267-311.
Savulescu, J., Hemsley, M., Newson, A., Foddy, B. 2006. Behavioural genetics: Why eugenic selection is preferable to enhancement. *Journal of Applied Philosophy* 23(2): 157-171.
Sidgwick, H. 1907 [1967]. *The methods of ethics* (7th ed.). London: Macmillan.
Temkin, L. S. 1987. Intransitivity and the mere addition paradox. *Philosophy and Public Affairs* 16(2): 138-187.
—. 1993a. *Inequality*. Oxford: Oxford University Press.
—. 1993b. Harmful goods, harmless bads. In *Value, welfare and morality*(eds.) R. G. Frey and C. W. Morris, 291- 324, Cambridge: Cambridge University Press.
Vallentyne, P. 1988. Gimmicky representations of moral theories. *Metaphilosophy* 19(34): , 253-63.

[1] Temkin (1993a, b) claims that this restriction, which he dubs "the Slogan", is presupposed in many arguments in moral philosophy, political theory, and welfare economics. The term "Person Affecting Restriction," introduced by Glover (1977), p. 66 (but see also Narveson (1967)), might be misleading since many theorists would, sensibly I think, lessen the restriction to also include other sentient beings. Cf. Holtug (1996). Below, I shall only discuss applications of the restriction to human populations. Consequently, whenever I claim that a certain interpretation of the restriction is reasonable, this claim only holds for human populations.

[2] See Parfit (1984), p. 388. For an overview of these counterintuitive implications, see Arrhenius et al (2006c) and Arrhenius (2000a).

[3] See Arrhenius (2000a, 2003b, 2006b), Broome (1992, 2004), Parfit (1984), Temkin (1993a, b).

[4] I've taken the term "uniquely realisable person" from Bykvist (1998) albeit his usage is slightly different from mine. There are a number of related but different views that appear in the literature (albeit seldom made explicit) that often are conflated with Comparativism and the Person Affecting Restriction. Comparativism should be distinguished from *Presentism* which draws a distinction between presently existing people and non-existing people; *Necessitarianism* which distinguishes between people who exist or will exist irrespective of how we act and people whose existence is contingent on our choices; and *Actualism* which differentiates people who have existed, exist or are going to exist in the actual world, on the one hand, and people who haven't, don't, and won't exist, on the other. These distinctions don't amount to the same thing but there are relations among them. A presently existing person is also a necessary and actual person but not the

other way around since necessary and actual people may be located in the past and the future. A necessary person is also an actual person but a future actual person may be contingent on our choice. A uniquely realisable person is also a contingent person, but a contingent person is not necessarily uniquely realisable in respect to all pairs of outcomes in a choice situation since she can exist, for instance, in two out of three outcomes. I discuss these distinctions at length in Arrhenius (2000a, 2006b).

[5] Perhaps it is this restriction which is at stake in Moore's criticism of Sidgwick at the turn of the century. It can be seen as a denial of Moore's idea in *Principia Ethica* that an unpopulated beautiful world is intrinsically better than an unpopulated ugly world, and a reaffirmation of Sidgwick's view that all moral goods must be of "Human Existence." See Moore (1903), section 50, and Sidgwick (1907), Bk. I, ch. IX, section 4.

[6] Cf., however, note 1 above.

[7] I have taken the term "personal good" from Broome (1991), ch. 8. The Personal Good Restriction is not, however, equivalent to his principle of personal good.

[8] See Arrhenius (2009) for a discussion of Welfarist Egalitarianism and population ethics. Broome (1991), pp. 180-1, suggests a way of understanding the goodness of equality that turns it into a personal good.

[9] An interesting question is whether the restriction should be supplemented with a necessary condition for outcomes being equally good. I would suggest the following condition: If outcome A is equally as good as B, then A and B are equally as good for at least one individual, or A is better (worse) and B is worse (better) for at least one individual. I'm grateful to Melinda Roberts for spotting an error in an earlier formulation of this principle.

[10] I'm here making the trivial assumption that if A is worse (better) than B for no one, then B is better (worse) than A for no one. Cf. note 21.

[11] Broome (1992), p. 124. Broome rejects the Person Affecting Restriction.

[12] Heyd (1988), p. 157. See also Heyd (1992), pp. 124-5. Cf. Narveson (1967), p. 67: "If you ask, 'whose happiness has been increased as a result of his being born?', the answer is that nobody's has. —Remember that the question we must ask about *him* is not whether he is happy but whether he is happier as a result of being born. And if put this way, we see that again we have a piece of nonsense on our hands if we suppose the answer is either 'yes' or 'no.' For if it is, then with whom, or with what, are we comparing his new state of bliss? Is the child, perhaps, happier than he used to be before he was born? Or happier than his alter ego? Obviously, there can be no sensible answer here." (emphasis in original).

[13] Heyd (1988), pp. 159 - 61 (emphasis in original). Heyd says that his view is "grounded in an 'anthropocentric' conception of value according to which value is necessarily related to human interests, welfare, expectations, desires and wishes—that is to say to human volitions" (p. 164). How this "volitional concept of value" is supposed to generate the conclusions quoted in the text is not clear to me and, as I pointed out earlier, the success of such deductions seems unlikely. Heyd also argues against Asymmetry by claiming that it "is inconsistent with a person-affecting theory as it presupposes the comparability of non-existence with life of a certain quality."

[14] Roberts (2003a), p. 1 (emphasis in original except for the last one). Notice that Roberts is here only talking about mere additions of extra people.

[15] Roberts (2003b), p. 179. I don't think, however, that Roberts is an Asymmetrical Strict Comparativist but rather, as I shall explain below, a Soft Comparativist.

[16] This case is basically a version of the Depletion case that Parfit (1984), pp. 361-2, uses as an argument against the Person Affecting Restriction.

[17] Broome (1999), ch. 10, p. 168 (emphasis in original). See also Parfit (1984), pp. 395 and 489, and Heyd (1988).

[18] For an effort to defuse Broome's argument, see Roberts (2003b). I shall say a bit more about Roberts' argument below.

[19] A similar example is used by Temkin (1987), pp. 168-9, to illustrate the intransitivity of the Person Affecting Restriction.

[20] For some sceptical arguments to this effect, see Bykvist (2007).

[21] I first suggested this in Arrhenius (2003b). I think that Wlodek Rabinowicz suggested something like this in personal communication already back in 2000 and he has put his take on it in print in Rabinowicz (forthcoming). For a recent incisive critique of the idea, see Bykvist (2007). Another possibility has been suggested by Holtug (1996). Again, the reason why some theorists have been inclined to deny that it can be better or worse for a person to exist than not to exist is that they believe that this position implies that it can be better or worse for a "person" not to exist than to exist and they find this implication nonsensical. One can deny this implication by revising the logic of "better for" and hold that it can be better or worse for a person to exist than not to exist, but it cannot be better or worse for a "person" not to exist than to exist. It is clear that a state X is better than a state Y if and only if state Y is worse than state X (I would say that this is a conceptual truth). Holtug denies that this logic also holds for "better for," that is, a state X is better for a person than another state Y if and only if state Y is worse for the person than state X, the reason being that "better for" is only applicable when a person to which the "for" in "better for" refers to exists. I consider Holtug's radical revision of the logic of "better for" an all

too high price to pay, especially in the light of my suggestion above which yields the same evaluations without any such revision.

[22] With one qualification: it involves a weak dominance condition which we didn't include in the description of the Personal Good Restriction.

[23] I discuss this latter claim at length in Arrhenius (2005).

[24] Following Vallentyne (1988), we could call a dilemma of the above mentioned type a "prohibition dilemma." There are also "obligation dilemmas," that is, situations where more than one action is obligatory. Since the normative version of Asymmetrical Comparativism is extensionally equivalent with Normative Strict Comparativism in cases that only involve lives with positive welfare, it also implies that Diagram 3 depicts a moral dilemma.

[25] Arrhenius (2005).

[26] Roberts (1998, 2003a).

[27] Roberts (2003a), pp. 10-13. See also Roberts (1998), pp. 62-65.

[28] Roberts (2003a), p. 10. See also Roberts (1998), p. 64. The word "wronged" has two meanings in English which are relevant in the present context (see e.g., Oxford English Dictionary). One meaning is more or less synonymous to "harmed" or "affected harmfully," another to "unjustly treated." I take it that Roberts is using the term in the second meaning, or something close to that.

[29] For example, Roberts (2003a), p. 11, where she discusses whether *r* is wronged in C.

[30] It seems trivially true that a person doesn't have any welfare level at all in a state in which she doesn't exist (well-being presupposes, so to say, being). A better solution, in my mind, is to revise P2 along the following lines:

A person p is not wronged at X if

(i) for each Y that is an alternative to X, W(*p*, X) ≥ W(*p*, Y); or
(ii) (ii) W(*p*, X) ≥ 0 and for each Y that is an alternative to X and in which *p* exists, W(*p*, X) ≥ W(*p*, Y).

[31] Roberts (2003b), p. 169. Moreover, Roberts (1998), p. 64, writes that "I am thus supposing that it is at least possible that *s* has more well-being in a world in which *s* does not exist than *s* actually has. Suppose *s*'s existence in X is unavoidably *less* than one worth living . . . and that *s* has, in any world in which *s* does not exist, a zero level of well-being. Under these conditions, *s*'s level of well-being at zero is actually *greater* than *s*'s well-being in X . . . . Thus, [P2] avoids the implication that the person whose life is less than one worth living has *not* been wronged" (emphasis in original). On the other hand, she also claims that "[t]here is no need, ever, on my account of either wrongful life or the non-identity problem to assign a value, *even a value of zero*, to nonexistence for an individual who never exists in the world that is subject to appraisal." (Roberts (1998), p. 174-5) (emphasis in original).

[32] In personal communication, Roberts endorses the same view.

[33] I haven't seen the person affecting version of Soft Comparativism explicitly stated in the literature but it seems to me that ideas somewhat along these lines are put forward in Hope (2003), Meyer (2003), and Roberts (1998, 2003a, 2007), and by Andrew Williams in private communication on the lawn of Worchester College summer 2003. At least I've been inspired by these sources to develop the above theory. Julian Savulescu has also proposed such a view in private communication. In Savulescu et al. (2006), p. 163, it is, however, stated that "[a]ccording to a person affecting view of harm, a person is harmed by an act if she is made worse off than she would otherwise have been if that act had not been performed." This sounds more like the counterfactual analysis of harm which is

plagued by a number of well-known problems. For a discussion of these in connection with population ethics, see Roberts (2007).

[34] Parfit (1984), p. 367. I've changed the wording of the example.

[35] Notice that the notion of "comparative harm" above is a technical notion that doesn't completely map onto our everyday use of "harm." For example, if you will enjoy an excellent life in both outcome A and B but you are slightly less happy in B, then you are comparatively harmed if B came about, but many would hesitate to say that you are harmed in the ordinary language sense of "harm" (there are many other examples). I could have used some other term to capture the idea that it makes an outcome worse if people are worse off than they otherwise could have been, but I think the technical notion of "comparative harm" is sufficiently related to the ordinary notion of harm to justify its name. An analogy would be the difference between the notion of the "consequence of an action" in ordinary language and in the formulation of consequentialism (usually the whole possible world that would be the case if the action were performed). For brevity, I will in the following sometimes use the term "harm" and its cognates although I always have in mind "comparative harm" in the above sense.

[36] Alternatively, we could represent the value of an outcome with an ordered pair (w, h) in which w represents the value of the well-being in the outcome and h represents the (negative) value of the harm in the outcome. Such a representation would leave open the possibility that comparative harm has lexical priority over total welfare in the ranking of outcomes. As I shall discuss below, such a view is not very attractive.

[37] Below, I'll assume that the harm factor is determined relative the optimal outcome for the people in question. There are of course other alternatives but I shall not pursue this question further here.

[38] Assume that the total well-being in C is m and in B thus $m + d$. The difference in value between B and C is then $(m + d - h) - m = d - h$.

[39] See Feldman (1997) and Arrhenius (2003a, 2006a, 2007).

[40] The kind of consequentialism I have in mind is what we could call Ceteris Paribus Act-Consequentialism: other things being equal, an action is right (obligatory) if and only if its outcome is at least as good as (better than) that of every alternative. An action is wrong if and only if it is not right. In other words, if a choice situation doesn't involve actions that are right or wrong by virtue of a certain deontic constraint or option, then the normative status of the actions are determined by the value of their respective outcomes. Most deontologists accept this form of consequentialism. For a discussion, see Arrhenius (2005).

[41] Strictly speaking, Soft Comparativism in conjunction with consequentialism has this implication. For the sake of brevity, I'll omit this qualification below. Moreover, I'm here assuming that probable outcomes are relevant for determining comparative harm.

[42] Another interesting possibility, suggested to me by an anonymous referee on another paper, is to claim that people are comparatively harmed if and only if they are worse off than they could have been both in terms of expected and actual welfare. According to this view, if we choose a_A and A actually came about, then the x-people are not harmed since their actual welfare is maximised. A problem for such a theory, however, is that its normative prescriptions are a bit unclear. Does it direct us to choose a_A or a_B?

[43] See, for example Carlson (1995) and Feldman (2006).

[44] I'm indebted to John Broome for pressing these two points.

[45] See Bales (1971) for an excellent treatment of this issue. See also Brink (1986), pp. 421-7, Danielsson (1974), pp. 28-9, Kymlicka (1990), p. 29.

[46] For the Repugnant Conclusion, see Parfit (1984), ch. 17. Cf. Arrhenius (2000a), ch. 3.

[47] One could also construct examples in which the y-people have hellish lives, like the case described in Diagram 1. However, such counterexamples could be avoided by revising Soft Comparativism such that it counts as being comparatively harmed if you are born into a life not worth living and there is an alternative in which you're not brought into existence.

[48] I would like to thank Julian Savulescu, Torbjörn Tännsjö, Dan Wikler, Andrew Williams, and especially John Broome, Krister Bykvist, Björn Eriksson, and Melinda Roberts, for fruitful discussions and comments. An earlier version of this paper was presented at the Högre Seminariet, Dept. of Philosophy, Stockholm University, October 2007. I would like to thank the participants in that seminar for their stimulating criticism. Thanks also to CERSES, CNRS, for being such a generous host during some of the time when this paper were written. Financial support from the Bank of Sweden Tercentenary Foundation, STINT, and the Swedish Collegium for Advanced Study is gratefully acknowledged. Some of the ideas discussed in this paper also appear in Arrhenius (2000a, 2003b, 2008).